

VIDEO SUMMARY DESCRIPTION SCHEME AND METHOD AND SYSTEM OF
VIDEO SUMMARY DESCRIPTION DATA GENERATION FOR EFFICIENT
OVERVIEW AND BROWSING

TECHNICAL FIELD

5 The present invention relates to a video summary description scheme for efficient video overview and browsing, and also relates to a method and system of video summary description generation to describe video summary according to the video summary description scheme.

10 The technical fields in which the present invention is involved are content based video indexing and browsing/searching and summarizing video to the content based and then describing it.

BACKGROUND OF THE INVENTION

15 The format of summarizing video largely falls into dynamic summary and static summary. The video description scheme according to the present invention is for efficiently describing the dynamic summary and the static summary into the unification based description scheme.

20 Generally, because the existing video summary and description scheme provide simply the information of video interval which is included in the video summary, the existing video summary and description scheme are limited to conveying overall video contents through the playing of the summary video.

However, in many cases, the browsing for identifying and revisiting concerned parts through overview of overall contents is needed rather than only overview of overall contents through the summary video.

25 Also, the existing video summary provides only the video interval which is considered to be important according to the criteria determined by the video summary provider. Accordingly, if the criteria of users and the video provider are different from

006260-18652960

each other or users have special criteria, the users can not obtain video summary of their desires.

That is, although the existing summary video permits the users selecting the summary video with desired level by providing several levels' summary videos, it makes the selecting extent of the users to be limited that the users can not select by the contents of the summary videos.

The US patent 5,821,945 entitled "Method and apparatus for video browsing based on content and structure" represents video in compact form and provides browsing functionality accessing to the video with desired content through the representation.

However, the patent is on the static summary based on the representative frame and although the existing static summary summarizes by using the representative frame of the video shot, the representative frame of this patent provides only visual information representing the shot, the patent has limitation on conveying the information using summary.

As compared with the patent, the video description scheme and browsing method utilize the dynamic summary based on the video segment.

The video summary description scheme was proposed by the MPEG-7 Description Scheme (V0.5) announced ISO/IEC JTC1/SC29/WG11 MPEG-7 Output Document No. N2844 on July 1999. Because the scheme describes the interval information of each video segment of dynamic summary video, in spite of providing basic functionalities describing dynamic summary, the scheme has problem in following aspects.

First, there is the drawback that it can not provide access to original video from summary segments constituting the summary video. That is, the users wanted to access to the original video to understand more detailed information on the basis of the summary contents and overview through summary video, however the existing scheme could not meet the need.

Secondly, the existing scheme can not provide sufficient audio summary description functionalities.

And finally, there is the drawback that in the case of representing event based summary, the duplicate description and the complexity of searching is indispensable.

SUMMARY OF THE INVENTION

An object of the present invention is to provide a hierarchical video summary description scheme, which comprises the representative frame information and the representative sound information at each video interval which is included in the summary video and makes the user customized event based summary providing users' selection for the contents of the summary video and efficient browsing to be feasible, and a video summary description data generation method and system using the description scheme.

In order to achieve the object, the HierarchicalSummary DS according to an executable example of the present invention comprises at least one HighlightLevel DS which is describing highlight level, and the HighlightLevel DS comprises at least HighlightSegment DS which is describing highlight segment information constituting the summary video of the highlight level.

Preferably, the HighlightLevel DS is composed of at least one lower level HighlightLevel DSs.

More preferably, the HighlightSegment DS comprises a VideoSegmentLocator DS which is describing time information or video itself of said corresponding highlight segment.

It is preferable that the HighlightSegment DS further comprises ImageLocator DS which is describing the representative frame of said corresponding highlight segment.

It is more preferable that the HighlightSegment DS further comprises SoundLocator DS which is describing the representative sound information of said corresponding highlight segment.

Preferably, the HighlightSegment DS further comprises ImageLocator DS which is describing the representative frame of said corresponding highlight segment and

SoundLocator DS which is describing the representative sound information of said corresponding highlight segment.

More preferably, the ImageLocator DS describes time information or image data of the representative frame of video interval corresponding to said corresponding
5 highlight segment.

Preferably, the HighlightSegment DS further comprises AudioSegmentLocator DS which is describing the audio segment information constituting an audio summary of said corresponding highlight segment.

More preferably, the AudioSegmentLocator DS describes time information
10 or audio data of the audio interval of said corresponding highlight segment.

It is preferable that the HierarchicalSummary DS includes SummaryComponentList describing and enumerating all of the SummaryComponentTypes which is included in the HierarchicalSummary DS.

Also, it is preferable that the HierarchicalSummary DS includes
15 SummaryThemeList DS which is enumerating the event or subject comprised in the summary and describing the ID and then describes event based summary and permits the users to browse the summary video by the event or subject described in said SummaryThemeList

It is more preferable that the SummaryThemeList DS includes arbitrary
20 number of SummaryThemes as elements and said SummaryTheme includes an attribute of id representing the corresponding event or subject, and the SummaryTheme further includes an attribute of parentID which is to describe the id of the event or subject of the upper level

Preferably, the HighlightLevel DS includes an attribute of themeIds
25 describing said attribute of ids of common events or subjects if all of the HighlightSegments and HighlightLevels which are constituting corresponding highlight level have common events or subjects.

More preferably, the HighlightSegment DS includes an attribute of themeIds describing said attribute of id and describes the event or subject of the corresponding highlight segment.

Also, according to the present invention, a computer-readable recording medium where a HierarchicalSummary DS is stored therein is provided. Preferably, the HierarchicalSummary DS comprises at least one HighlightLevel DS which is describing highlight level, and the HighlightLevel DS comprises at least one HighlightSegment DS which is describing highlight segment information constituting the summary video of that the highlight level, and the HighlightSegment DS comprises VideoSegmentLocator DS describing time information or video itself of said corresponding highlight segment.

Also, according to the present invention, a method for generating video summary description data according to video summary description scheme by inputting original video is provided. The method includes the following steps: video analyzing step which is producing video analysis result by inputting the original video and then analyzing the original video; summary rule defining step which is defining the summary rule for selecting summary video interval; summary video interval selecting step which is constituting summary video interval information by selecting the video interval capable of summarizing video contents from the original video by inputting said original video analysis result and said summary rule; and video summary describing step which is producing video summary description data according to the HierarchicalSummary DS by inputting the summary video interval information output by said summary video interval selecting step.

Preferably, the video analyzing step comprises feature extracting step which is outputting the types of features and video time interval at which those features are detected by inputting the original video and extracting those features, event detecting step which is detecting key events included in the original video by inputting said types of features and video time interval at which those features are detected; and episode detecting step which is detecting episode by dividing the original video into story flow base on the basis of said detected event:

Preferably, the summary rule defining step provides the types of summary events, which are bases in selecting the summary video interval, after defining them to said video summary describing step.

5 More preferably, the method further comprises representative frame extracting step which is providing the representative frame to said video summary describing step by inputting said summary video interval information and extracting representative frame.

10 More preferably, the method further comprises representative sound extracting step which is providing the representative sound to said video summary describing step by inputting said summary video interval information and extracting representative sound.

15 Also, according to the present invention, a computer-readable recording medium where a program is stored therein is provided. The program executes the following steps: feature extracting step which is outputting the types of features and video time interval at which those features are detected; event detecting step which is detecting key events included in the original video by inputting said types of features and said video time interval at which those features are detected; episode detecting step which is detecting episode by dividing the original video into story flow base on the basis of said detected key events; summary rule defining step which is defining the summary rule for selecting the
20 summary video interval; summary video interval selecting step which is constituting summary video interval information by selecting the video interval capable of summarizing the video contents of the original video by inputting said detected episode and said summary rule; and video summary describing step which is generating video summary description data with HierarchicalSummary DS by inputting the summary video interval
25 information output by said summary video interval selecting step.

Also, according to the present invention, a system for generating video summary description data according to video summary description scheme by inputting original video is provided. The system includes video analyzing means for outputting video analysis result by inputting original video and analyzing the original video, summary rule

means by inputting said summary video interval information and extracting representative sound.

Also, according to the present invention, a computer-readable recording medium where a program is stored therein is provided. The program is for functioning
5 feature extracting means for outputting the types of features and video time interval at which those features are detected, event detecting means for detecting key events included in the original video by inputting said types of features and said video time interval at which those features are detected, episode detecting means for detecting episode by dividing the original video into story flow base on the basis of said detected key events,
10 summary rule defining means for defining the summary rule for selecting the summary video interval, summary video interval selecting means for constituting summary video interval information by selecting the video interval capable of summarizing the video contents of the original video by inputting said detected episode and said summary rule, and video summary describing means for generating video summary description data with
15 HierarchicalSummary DS by inputting the summary video interval information output by said summary video interval selecting step.

Also, a Video browsing system in a server/client circumstance according to the present invention is provided. The system includes a server which is equipped with video summary description data generation system which generates video summary
20 description data on the basis of HierarchicalSummary DS by inputting original video and links said original video and video summary description data, and a client which is browsing and navigating video by overview of said original video and access to the original video of said server using said video summary description data.

BRIEF DESCRIPTION OF THE DRAWINGS

25 The embodiments of the present invention will be explained with reference to the accompanying drawings, in which:

FIG. 1 is a block diagram illustrating a system for generating video summary description data according to the description scheme of the present invention.

FIG. 2 is a drawing that illustrates the data structure of the HierarchicalSummary DS describing the video summary description scheme according to the present invention in UML (Unified Modeling Language).

FIG. 3 is a compositional drawing of user interface of the tool for playing and browsing of the summary video inputting the video summary description data described by the same description scheme as FIG. 2.

FIG. 4 is a compositional drawing for the flow of the data and control for hierarchical browsing using the summary video of the present invention.

DETAILED DESCRIPTION OF THE INVENTION

10 The present invention will be described in detail by way of a preferred embodiment with reference to accompanying drawings, in which like reference numerals are used to identify the same or similar parts.

FIG. 1 is a block diagram illustrating a system for generating video summary description data according to the description scheme of the present invention.

15 As illustrated in FIG. 1, the apparatus for generating video description data according to the present invention is composed of a feature extracting part 101, an event detecting part 102, an episode detecting part 103, a summary video interval selecting part 104, a summary rule defining part 105, a representative frame extracting part 106, a representative sound extracting part 107 and a video summary describing part 108.

20 The feature extracting part 101 extracts necessary features to generate summary video by inputting the original video. The general features include shot boundary, camera motion, caption region, face region and so on.

In the step of extracting features, the types of features and video time interval at which those features are detected are output to the step of detecting event in the format of (types of features, feature serial number, time interval) by extracting those features.

25

For example, in the case of camera motion, (camera zoom, 1, 100 ~ 150) represents the information that the first zoom of camera was detected in the 100 ~ 150 frame.

The event detecting part 102 detects key events which are included in the original video. Because these events must represent the contents of the original video well and are the references for generating summary video, these events are generally differently defined according to genre of the original video.

These events either may represent higher meaning level or may be visual features which can directly infer higher meaning. For example, in the case of soccer video, goal, shoot, caption, replay and so on can be defined as events.

The event detecting part 102 outputs the types of detected events and the time interval in the format of (types of events, event serial number, time interval). For example, the event information indicating that the first goal occurred at between 200 and 300 frame is output in the format of (goal, 1, 200 ~ 300).

The episode detecting part 103, on the basis of the detected event, divides the video into an episode with larger unit than an event based on the story flow. After detecting key events, an episode is detected while including accompanied events which follow the key event. For example, in the case of soccer video, goal and shoot can be key events and bench scene, audiences scene, goal ceremony scene, replay of goal scene and so on compose accompanied events of the key events.

That is, the episode is detected on the basis of the goal and shoot.

The episode detection information is output in the format of (episode number, time interval, priority, feature shot, associated event information). Herein, the episode number is serial number of the episode and the time interval represents the time interval of the episode by the shot unit. The priority represents the degree of importance of the episode. The feature shot represents the shot number including the most important information out of the shots comprising the episode and the associated event information represents the event number of the event related to the episode. For example, in the case of representing the episode detection information as (episode 1, 4 ~ 6, 1, 5, goal 1, caption 3),

the information means that the first episode includes 4 ~ 6th shot, the priority is the highest (1), the feature shot is fifth shot, and the associated events are the first goal and the third caption.

The summary video interval selecting part 104 selects the video interval at which the contents of the original video can be summarized well on the basis of the detected episode. The reference of selecting the interval is performed by the predefined summary rule of the summary rule defining part 105.

The summary rule defining part 105 defines rule for selecting the summary interval and outputs control signal for selecting the summary interval. The summary rule defining part 105 also outputs the types of summary events, which are bases in selecting the summary video interval, to the video summary describing part 108.

The summary video interval selecting part 104 outputs the time information of the selected summary video intervals by frame units and outputs the types of events corresponding to the video intervals. That is, the format of (100 ~ 200, goal), (500 ~ 700, shoot) and so on represent that the video segments selected as the summary video intervals are 100 ~ 200 frame, 500 ~ 700 frame and so on and the event of each segment is goal and shoot respectively. As well, the information such as file name can be output to facilitate the access of an additional video which is composed of only the summary video interval.

If the summary video interval selection is completed, the representative frame and the representative sound are extracted from the representative frame extracting part 106 and the representative sound extracting part 107 respectively by using the summary video interval information.

The representative frame extracting part 106 outputs the image frame number representing the summary video interval or outputs the image data.

The representative sound extracting part 107 outputs the sound data representing the summary video interval or outputs the sound time interval.

The video summary describing part 108 describes the related information in order to make efficient summary and browsing functionalities to be feasible according to the Hierarchical Summary Description Scheme of the present invention shown in FIG. 2.

The main information of the Hierarchical Summary Description Scheme comprises the types of summary events of the summary video, the time information describing each summary video interval, the representative frame, the representative sound, and the event types in each interval.

5 The video summary describing part 108 outputs the video summary description data according to the description scheme illustrated in FIG. 2.

FIG. 2 is a drawing that illustrates the data structure of the HierarchicalSummary DS describing the video summary description scheme according to the present invention in UML (Unified Modeling Language).

10 The HierarchicalSummary DS 201 describing the video summary is composed of one or more HighlightLevel DS 202 and one or zero SummaryThemeList DS 203.

15 The SummaryThemeList DS provides the functionality of the event based summary and browsing by enumeratively describing the information of subject or event constituting the summary. The HighlightLevel DS 202 is composed of the HighlightSegment DSs 204 as many as the number of the video intervals constituting the summary video of that level and zero or several number of HighlightLevel DS.

20 The HighlightSegment DS describes the information corresponding to the interval of each summary video. The HighlightSegment DS is composed of one VideoSegmentLocator DS 205, zero or several ImageLocator DSs 206, zero or several SoundLocator DSs 207 and AudioSegmentLocator 208.

The followings give more detailed description about the HierarchicalSummary DS.

25 The HierarchicalSummary DS has an attribute of SummaryComponentList which obviously represents the summary type, which is comprised by the HierarchicalSummary DS.

The SummaryComponentList is derived on the basis of the SummaryComponentType and describes by enumerating all comprised SummaryComponentTypes.

006260-1857960 09675984-092900

In the SummaryComponentList, there are five types such as keyFrames, keyVideoClips, keyAudioClips, keyEvents, and unconstraint.

The keyFrames represents the key frame summary composed of representative frames. The keyVideoClips represents the key video clip summary composed of key video intervals' sets. The keyEvents represents the summary composed of the video interval corresponding to either the event or the subject. The keyAudioClips represents the key audio clip summary composed of representative audio intervals' sets. And, the unconstraint represents the types of summary defined by users except for said summaries.

Also, in order to describe event based summary, the HierarchicalSummary DS might comprise the SummaryThemeList DS which is enumerating the event (or subject) comprised in the summary and describing the ID.

The SummaryThemeList has arbitrary number of SummaryThemes as elements. The SummaryTheme has an attribute of id of ID type and selectively has an attribute of parentId.

The SummaryThemeList DS permits the users browsing the summary video from the viewpoint of each event or several subjects described in the SummaryThemeList. That is, the application tool inputting description data makes the users to select the desired subject by parsing the SummaryThemeList DS and providing the information to the users.

At this time, in the case of enumerating these subjects into simple format, if the number of the subjects are large, it might not easy to find out the subject desired by the users.

Accordingly, by representing the subject as tree structure similar to ToC (Table of Content), the users efficiently can do browsing at each subject after finding out the desired subject.

In order to do so, the present invention permits the attribute of parentId being selectively used in the SummaryTheme. The parentId means the upper element (upper subject) in the tree structure.

006260"48657960

The HierarchicalSummary DS of the present invention comprises HighlightLevel DSs and each HighlightLevel DS comprises one or more HighlightSegment DS which corresponds to a video segment (or interval) constituting the summary video.

The HighlightLevel DS has an attribute of themeIds of IDREFS type.

5 The themeIds describes the subject and event id, common to the children HighlightLevel DS of corresponding HighlightLevel DS or all HighlightSegment DSs comprised in the HighlightLevel, and the id is described in said SummaryThemeList DS.

10 The themeIds can denote several events and, when doing event based summary, solve the problem that same id is unnecessarily repeated in all segments constituting the level by having the themeIds representing common subject type in the HighlightSegment constituting the level.

The HighlightSegment DS comprises one VideoSegmentLocator DS and one or more ImageLocator DS, zero or one SoundLocator DS and zero or one AudioSegmentLocator DS.

15 Herein, the VideoSegmentLocator DS describes the time information or video itself of the video segment constituting the summary video. The ImageLocator DS describes the image data information of the representative frame of the video segment. The SoundLocator DS describes the sound information representing the corresponding video segment interval. The AudioSegmentLocator DS describes the interval time information of
20 the audio segment constituting the audio summary or the audio information itself.

The HighlightSegment DS has an attribute of themeIds. The themeIds describes using the id defined in the SummaryThemeList which subjects or events described in said SummaryThemeList DS relates to the corresponding highlight segment.

25 The themeIds can denote more than one events and by allowing one highlight segment to have several subjects, it is an efficient technique of the present invention which is solving the problem of indispensable duplication of descriptions caused by describing the video segment at each event (or subject) when using the existing method for event based summary.

The user accesses to the original video shot corresponding to the representative frame by clicking the representative frame.

5 The summary video level 0 representative frame part and the representative sound part 307 and the summary video level 1 representative frame part and the representative sound part 306 shows the frame and sound information representing each video interval of the summary video level 0 and the summary video level 1 respectively. That is, it is composed of the iconic images representing a series of the images and sounds with reduced sizes.

10 If the user clicks the representative frame of the summary video representative frame part and the representative sound part, the user accesses to the original video interval corresponding to the representative frame. Herein, in the case of clicking the representative sound icon corresponding to the representative frame of the summary video, the representative sound of the video interval is played.

15 The summary video controlling part 302 inputs the control for user selection to play the summary video. In the case of being provided with the multi level summary video, the user does overview and browsing by selecting the summary of the desired level through the level selecting part 303. The event selecting part 304 enumerates the event and the subject provided by the SummaryThemeList and the user does overview and browsing by selecting the desired event. After all, this realizes the summary of the user customization
20 type.

FIG. 4 is a compositional drawing for the flow of the data and control for hierarchical browsing using the summary video of the present invention.

25 The browsing is performed by accessing the data for browsing with the method of FIG. 4 through the use of the user interface of FIG.3. The data for browsing are the summary video and the representative frame of the summary video and the original video 406 and the original video representative frame 405.

The summary video is assumed to have two levels. Needless to say, the summary video may have more levels than two. The summary video level 0 401 is what is summarized with shorter time than the summary video level 1 403. That is, the summary

video level 1 contains more contents than the summary video level 0. The summary video level 0 representative frame 402 is the representative frame of the summary video level 0 and the summary video level 1 representative frame 404 is the representative frame of the summary video level 1.

5 The summary video and the original video are played through the video playing part 301 of FIG. 3. The summary video level 0 representative frame is displayed in the summary video level 0 representative frame and the representative sound part 306, the summary video level 1 representative frame is displayed in the summary video level 1 representative frame and the representative sound part 307, and the original video
10 representative frame is displayed in the original video representative frame part 305.

 The hierarchical browsing method illustrated in FIG. 4 can have various types of hierarchical paths as the following example.

Case 1 : (1) – (2)

Case 2 : (1) – (3) – (5)

Case 3 : (1) – (3) – (4) – (6)

Case 4 : (7) – (5)

Case 5 : (7) – (4) – (6)

The overall browsing scheme is as follows.

First, understand the overall contents of the original video by watching the
20 summary video of the original video. Herein, the summary video may play either the summary video level 0 or the summary video level 1. When more detailed browsing is wanted after watching the summary video, the interested video interval is identified through the summary video representative frame. If the scene which is desired to be exactly found, is identified in the summary video representative frame, play it by directly accessing
25 to the video interval of the original video to which the representative frame is connected. And if the more detailed information is needed, the user may access to the desired original video either by understanding the representative frame of the next level or by hierarchically understanding the contents of the representative frame of the original video.

Although these hierarchical browsing techniques might take long time in browsing to access to the desired contents while the original video is being played, the browsing time is drastically reduced by directly accessing to the contents of the original video through the hierarchical representative frame.

- 5 The existing general video indexing and browsing techniques divide the original video in shot unit and access to the shot by perceiving the desired shot from the representative frame after constituting the representative frame representing each shot.

10 In this case, because the number of the shots of the original video is large, lots of time and efforts are necessary to do browsing the desired contents out of many representative frames.

In the present invention, it is feasible to quickly access to the desired video by constituting the hierarchical representative frame with the representative frame of the summary video.

- 15 The case 1 is the case that plays the summary video level 0 and directly accesses to the original video from the summary video level 0 representative frame.

20 The case 2 is the case that plays the summary video level 0 and selects the most interested representative frame from the summary video level 0 representative frame and identifies the desired scene in the summary video level 1 representative frame corresponding to the neighborhood of the representative frame to understand more detailed information before access to the original video and then accesses to the original video.

25 The case 3 is the case that selects the most interested representative frame to obtain more detailed information in the case that the access from the summary video level 1 representative frame to the original video is difficult in the case 2 and by the original video representative frames neighboring the representative frame identifies the desired scene and then accesses to the original video using the representative frame of the original frame.

The case 4 and case 5 are the cases that start at the playing of the summary video level 1 and the paths are similar to the above cases.

When applied to the server/client circumstance, the present invention can provide the system in which multiple clients access to one server and can do video overview and browsing. The original video is inputted to the server and the video summary description data is produced on the basis of the hierarchical summary description scheme and the summary video description data generation system linking said original video and the video summary description data is equipped. The client accesses to the server through the communication network, does overview of the video using the video summary description data and does browsing and navigation of the video by accessing to the original video.

Although, the present invention was described on the basis of preferably executable examples, these executable examples do not limit the present invention but exemplify. Also, it will be appreciated by those skilled in the art that changes and variations in the embodiments herein can be made without departing from the spirit and scope of the present invention as defined by the following claims.